MSDS 413, Summer 2019, Assignment 10
Non-linear Modeling and Model Monitoring (TS10)

Introduction

For this assignment, you will use the datasets DTload.csv, AustralianWines.csv, BullBear.csv, and BTC-USD.csv, which are included in the Session 10 zip file (TS10.zip) and, along with a R script (TS10.R), is posted to Canvas. You will read the files into R or other statistics package and conduct the requested analyses.

The following list defines the variables:

- DTload.csv
  - Date
  - Value
- AustralianWine.csv (volumes are 1000× Kiloliters
  - Month
  - Fortified
  - Red
  - Rose
  - sparkling
  - Sweet.white
  - Dry.white
- BullBear.csv
  - returns
  - actual
- S&P500 data from 2004 on
  - GSPC.Open
  - GSPC.High
  - GSPC.Low
  - GSPC.Close
  - GSPC.Volume
  - GSPC.Adjusted
- BTC-USD.csv

- Date
- Open
- High
- Low
- Close
- Adj.Close
- Volume

Your objective is to explore the time series behavior of these data sets including EDA, modeling, model diagnostics, and interpretation.

Procedure

The following steps are necessary to complete this assignment. Address each and every part and ensure that you cover all the details specified in the questions. Be sure to use all your analytical skills. Note: this is an exploratory assignment into nonlinear time series forecast methods. It is less rigorous than the previous assignments, so feel free to have a bit of fun. Do, however, keep on your analyst hat and comment accordingly.

1. **Random Forest** (2 points) Use the DTload data to forecast changes in the load. DTLoad is a time series representing electricity usage. Data are sampled every 30 minutes. There are two seasonal periods: daily (period=48) and weekly (period=336).
   1.1. Perform EDA.
   1.2. Build a random forest model specifying a one day forecast (the shortest season).
   1.3. Interpret what is shown in the forecast plot.
   1.4. Zoom in on the forecasts and interpret the forecasts behavior.
2. **Neural Network** (2 points) Use the Australian Wine data to forecast fortified wine sales using a neural network. Do the Problems (Section 9.7) on page 201, Shmueli and Lichtendahl (2018) as reproduced here.
   2.1. Use the EDA from the earlier assignment that used the Australian wine data.
   2.2. Partition the data using the period until December 1993 as the training period. Run a neural network using R's **nnetar** function with 11 non-seasonal lags (i.e., $p = 11$). Leave all other arguments at their default.
   2.3. Create a time plot for the actual and forecasted series over the training period. Create also a time plot of the forecast error for the training period. Interpret what you see in the plots.
   2.4. Use the nerural network to forecast sales for each month in the validation period (January 1994 to December 1994).

2.5. Compare your neural network to an exponential smoothing model used to forecast fortified wine sales using R's textbfets function to automatically select and fit an exponential smoothing model to the training period until December 1993. Which model did **ets** fit?

2.6. How does the neural network compare to the exponential smoothing model in terms of predictive performance in the training period?

2.7. How does the neural network forecast compare with some of the other models you constructed in Assignment TS1?

3. **HMM** (2 points) Use the BullBear and the S&P500 data to identify regime changes in these data.

3.1. EDA, of course.

3.2. Fit a Hidden Markov Model. Interpret the output.

3.3. Obtain the true regimes and the posterior probabilities of the regimes. Interpret.

3.4. Fit S&P500 returns with a two state HMM. Plot the returns and posterior probabilities. Interpret.

3.5. Fit a 3-state HMM to the S&P500 returns. Plot the returns stream and the posterior probabilities of the separate regimes. Interpret.

4. **SPRT** (1.5 points) Use the BTC-USD data to find exceptions.

4.1. Choose either natural, differenced, logged, or differenced logged closing prices. Justify your choice.

4.2. Use your choice of response to monitor the cryptocurrency time series iteratively to find where exceptions in the series occur.

4.3. Construct a ARMA or volatility model or both and use the SPRT code on the residuals.

4.4. Compare exception locations between parts 4.2. and parts 4.3.. Do you think one is superior to the other? Why?

5. **Compare** (1.5 points) Comment on what you think are the differences between the linear models of the first nine weeks with these nonlinear models of this week.

Deliverables

See Section Submission Directions below. The assignment deliverables, each in pdf format, are as follows:

- *Only if requested by instructor*
  - The program or script
  - Logs

  - Outputs

- Mandatory
  Data analysis write-up: no programs, logs, or just code outputs.

The data analysis must follow and use the item numbering of each assignment, i.e., use the numbers, say, 1 - 5, with the sub-lettering if used. These deliverables are provided according to the instructions in the Submission Directions section below.

Submission Directions

Title Page

Include a title page with your name and the assignment designation. Leave room for instructor comments.

File Names

The assignment write-up file shall be submitted to Canvas according to the schedule in the syllabus using the item (1) naming convention below. The naming convention is case sensitive. Use letters and numbers as given. **The file name parts have no spaces or other separator charancters**. TS10Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS10; your lastname with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS10Lastname.pdf

For example: Suppose your name is Student McStats. Your filename then is:

TS10Mcstats.pdf

The analysis write-up file must be submitted for grading. Each write-up requires a title page for instructor comments. The analysis may use either R or any other statistics package you wish, or if you use more than one package, you must use the germane tables, plots, etc., in a single report. If you use more than one package, differences and similarities should be indicated.

email: jamie.riggs@northwestern.edu

Email *ONLY IF REQUESTED* the program (script), log and output as separate pdf files. The R log and output may be combined. The file names shall be as follows:

- The program or script file names

  - TS10LastnameRprog.pdf

- The log file names
  - TS10LastnameRlog.pdf

References

G. Shmueli and K.C. JR. Lichtendahl. *Practical Time Series Forecasting with R: A Hands-On Guide*. Axelrod Schnall Publishers, 2 edition, 2018.